Intra-specific diversity in noncoding regions of the mitochondrial DNA of *Globodera pallida*

Nikola Grujić

European Master of Science in Nematology

Faculty of Science, Gent University
Gent, Belgium
June 2010

Supervisor: Dr Vivian Blok
Promoter: Dr John Jones
Intra-specific diversity in noncoding regions of the mitochondrial DNA of *Globodera pallida*

NIKOLA GRUJIC $^{1,2}$

$^1$Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, UK
$^2$Faculty of Science, Ghent University, K.L. Ledeganckstraat 35, 9000 Gent, Belgium

**Summary** - Potato cyst nematodes (PCN) *Globodera rostochiensis* and *G. pallida* are economically important pests of potato (*Solanum tuberosum* L.) and both PCN species are listed as quarantine pests within Europe. In addition these nematode species have a unique multipartite structure to their mitochondrial DNA. The aim of this study was to utilize this unusual feature to examine whether it could be used to differentiate and group populations of *G. pallida* which could then be used to identify any unusual populations which may have different virulence characteristics. Forty one populations of *G. pallida* from South America and Europe together with five populations of *G. rostochiensis* were used. Two areas of the mtDNA genome were examined. The main area examined was variation a part of the large non-coding “222” region found in scmtDNA II and scmtDNA IV. This was then compared the degree of variation found in the Cytochrome B gene. The “short222” region was amplified and then subjected to RFLP analysis using TaqI, AluI and XbaI restriction enzymes. RFLPs were visualised on agrarose gels and were compared to terminal fragment, T-RFLPs using fluorescently labeled primers. In addition a sub-sample of the *G. pallida* populations was sequenced to enable an in-silico comparison of digestions with T-RFLPs and an examination of the composition of scmtII and IV within individual nematodes. All the techniques showed differentiation of the populations with the main split being between the majority of the European populations and those from South America. Examination of the sequences from individual nematodes showed more than one type of “222” sequence could be present.

**Key words** – *Globodera pallida*, CytB, scmtDNA, RFLP, T-RFLP, phylogeny, population genetics, diagnostic

Corresponding author, e-mail: n.grujic@ugent.be, nikola.grujic@scri.ac.uk
Introduction

The root knot and cyst nematodes are amongst the most important agricultural pests world-wide with annual losses estimated to be in excess of 100 billion US Dollars. Of the cyst nematodes, the potato cyst nematodes (PCN) *Globodera rostochiensis* (Woll.) Skarbilovich, and *G. pallida* Stone are the most economically important pests of potato (*Solanum tuberosum* L.) and both PCN species are listed as EPPO A2 Annex 1/A2 quarantine pests. They are wide-spread in Europe and despite quarantine measures and EU directives, *G. rostochiensis* is now found in all EU countries that officially test to species level and *G. pallida* has been identified in most EU member states, although their incidence varies considerably across the EU. Despite legislative regulation in over 55 countries they continue to be found in new locations with recent detections in Idaho USA, Quebec Canada, Victoria Australia, Java Indonesia and the Ukraine indicating that new introductions still occur despite regulatory actions in place to prevent their spread. The recent detection of PCN in important potato producing regions around the world emphasizes the need to develop tools for early detection and to improve methods to characterize populations so that appropriate eradication, control and management strategies crops can be implemented.

In the UK, PCN are the most economically important nematode pests of agriculture. In a survey of England and Wales Minnis et al., (2002) found 65% of potato land positive for PCN. For UK yield losses of around 9% have been estimated (Haydock and Evans, 1998) and there is an annual cost in excess of £50 million and this threatens the economic viability of potato cultivation for some growers. In the UK rotations of six years are recommended to allow populations levels to decline between potato crops. Use of the *H1* resistance gene has been very effective and durable against *G. rostochiensis* but, in the UK this has lead to selection for *G. pallida* which is now present at 92% of sites in England and Wales with PCN.
(Minnis et al., 2002). The increased prevalence of *G. pallida* is predicted to occur in many other potato growing regions in the world where the H1 gene is employed to control *G. rostochiensis*. Future deployment of cultivars which are now becoming available that have high levels of resistance to some *G. pallida* pathotypes might lead to the increase in prevalence of other pathotypes which currently are below detection levels or localized in occurrence. The withdrawal in 2007 by the EU, of the nematicides such as aldicarb for use in controlling PCN on potato crops due to their potential effects on non-target organisms, highlights the need for effective methods to monitor and manage PCN.

Recent achievements in molecular biology and the wide application of molecular techniques have provided new tools for identification and quantification of plant parasitic nematodes (Blok, 2005) and increased our knowledge in the fields of taxonomy and phylogeny of the nematodes. Mitochondrial (mt) genomes have provided a rich source of genetic markers for population genetic studies and the increasing availability of mt genome sequence data provides a basis for studies of the physiology, biochemistry and molecular biology of the mitochondria of nematodes (Moritz et al., (1987); Lemire (2005); Tsang and Lemire (2003)). Several mt genes have been used as molecular markers in phylogenetic studies, however, heteroplasmacy (Tsang and Lemire (2002b) and recombination between different sequence types (Lunt and Hyman 1997, Armstrong et al., (2007)) complicates the use of mtDNA for these studies. For example, Plantard et al. (2008) used partial cytochrome b (CytB) sequences of potato cyst nematodes from Peruvian Andes to investigate the origin of European populations of *Globodera pallida* transferred from South America to Europe more than a century ago. These authors identified five clades within Peruvian populations of PCN corresponding to five geographic regions ranging from the south to the north of Peru, and all European populations of *G. pallida* clustered in first clade from southern region around
Titicaca Lake, which indicates introductions to Europe are probably from this particular region. Within clade I four sub-clusters were found. Similarly, Pylypenko et al. (2008) carried out a study comparing two populations of *G. pallida* from Ukraine with the other European and South American populations using cytochrome b gene partial sequence. They found that Ukrainian populations were identical to each other and to a number of European populations, and they all clustered in first clade from the south of Peru, as in Plantard et al. (2008).

Recent studies by Armstrong et al. (2000, 2007) and Gibson et al. (2007a, b) on the unique multipartite structure of the mtDNA for *Globodera* spp. have provided insights into genome organization and population genetics within these species. The mitochondrial genome of *Globodera pallida* consists of at least six small circular mitochondrial DNAs (scmtDNAs). They range in the size from ~6.3 to 9.5 kb which make them considerably smaller than the smallest nematode genome sequenced to-date and not big enough to encode all mt proteins and RNAs usually present on metazoan mtDNA (Armstrong et al., 2000). Presence of these scmtDNAs has been confirmed by sequence analysis (Armstrong et al. (2007), Gibson et al. (2007a)) and a molecule larger than 9.5kb has not been found to-date. Armstrong et al. (2000) and Gibson et al. (2007a) completely sequenced and mapped the genes onto each circle, denoted as scmtDNAs I-V, and they reported observations about the distribution of the scmtDNAs in populations of *G. pallida*, the occurrence and distribution of sequence variants of scmtDNA IV in population P4A from Peru and they also found evidence of recombination between sequence variants of scmtDNAs IV in P4A which gives rise to mosaic structures.

Armstrong et al. (2007) investigated the genetic relationships of four populations of *G. pallida*. Population Gourdie from Scotland is part of a large group of European *G. pallida*
Pa2/3 populations, which has been previously determined as very heterogeneous in terms of virulence and genomic structure (Blok et al., 1997). Population Luffness from Scotland is representative of the highly virulent Pa2/3 populations and it is molecularly distinct from most European populations. Two South American populations used in this study, P4A and P5A are molecularly dissimilar from each other and from European populations (Blok et al., 1997) and show differences in virulence (Phillips and Trudgill 1998). Armstrong et al. (2007) showed that population Gourdie contains scmtDNA II and scmtDNA III as the dominant mitotypes, whereas scmtDNA I is the dominant mitotype in population Luffness, confirming the previous results of Armstrong et al. (2000). The hybridization patterns observed for population P4A showed that it contains mixtures of scmtDNAs found in populations Gourdie and Luffness as well as size variants of scmtDNA IV, with at least three forms (Armstrong et al., 2007). These authors suggested that the predomination of scmtDNA II and III in Gourdie and scmtDNA I in Luffness indicated that scmtDNA lineages have the ability to segregate and be retained during the separation of populations. The distribution of variants of scmtDNA IV between cysts of P4A was further investigated by Armstrong et al. (2007). They sequenced 11 clones of PCR products generated from the large noncoding “222” region and part of the 16s ribosomal RNA of scmtDNA IV (~3.2kb) from single individuals of P4A. The aligned sequences gave three clades and it was concluded that individuals contain just one form of scmtDNA IV, however multiple forms were present in population.

The aim of this work was to exploit the unusual structure and diversity of the mtDNA of *G. pallida* to differentiate and group populations within the species and to identify populations with unusual mtDNA profiles. Early detection of unusual population which may have different virulence characteristics than the prevalent population is important for timely management of *G. pallida*. 
Materials and methods

NEMATODE POPULATIONS

38 populations of *G. pallida* and 6 populations of *G. rostochiensis* from Europe, South and North America were used in this study. Most of the populations were obtained from the SCRI PCN collection. The remaining populations were obtained from the SASA, UK PCN collection, Faculty of Agriculture, Serbia and 4 field populations from Luffness obtained from fields sampled in 2009 and 2010 in the vicinity of the original collection of the Luffness population, which showed molecular and virulence differences compared to other European *G. pallida* populations. Population names, their origin and source of sequence information are presented in Table 1.

**Table 1.** List of populations used in present study, their origin and source of s222 and CytB sequences.

<table>
<thead>
<tr>
<th>Population</th>
<th>Country of Origin</th>
<th>Species</th>
<th>Sequence source</th>
<th>S222</th>
<th>CytB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lindley*#</td>
<td>England</td>
<td><em>G. pallida</em></td>
<td>SCRI/this study</td>
<td>SCRI unpublished</td>
</tr>
<tr>
<td>2</td>
<td>Farcet*</td>
<td>England</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>SCRI unpublished</td>
</tr>
<tr>
<td>3</td>
<td>Newton*</td>
<td>England</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Fail to sequence</td>
</tr>
<tr>
<td>4</td>
<td>Bedale*</td>
<td>England</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>SCRI unpublished</td>
</tr>
<tr>
<td>5</td>
<td>Gourdie*#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>SCRI/this study</td>
<td>SCRI unpublished</td>
</tr>
<tr>
<td>6</td>
<td>Latch*</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>7</td>
<td>Windyedge*</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Fail to sequence</td>
</tr>
<tr>
<td>8</td>
<td>Cotton of Ovenstone*#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>9</td>
<td>Bankhead of Buchel*#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>This study results</td>
<td>Fail to sequence</td>
</tr>
<tr>
<td>10</td>
<td>Dothan*</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>11</td>
<td>Muirhouse95*#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>This study results</td>
<td>Fail to sequence</td>
</tr>
<tr>
<td>12</td>
<td>Blackhall*</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>13</td>
<td>Luffness 09*#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>This study results</td>
<td>This study results</td>
</tr>
<tr>
<td>14</td>
<td>BBA1*</td>
<td>Germany</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>15</td>
<td>BBA2*</td>
<td>Germany</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>16</td>
<td>Kalle*</td>
<td>Germany</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>17</td>
<td>Chavonery*</td>
<td>Switzerland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>EF189180.1</td>
</tr>
<tr>
<td>18</td>
<td>Rookmaker*</td>
<td>Netherlands</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>19</td>
<td>D375*</td>
<td>Netherlands</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>20</td>
<td>GP-74-768-20*</td>
<td>Netherlands</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>21</td>
<td>Guielain*</td>
<td>France</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>22</td>
<td>St Malo*</td>
<td>France</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>23</td>
<td>Uzhok*</td>
<td>Ukraine</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Fail to sequence</td>
</tr>
<tr>
<td>24</td>
<td>Zhornava*</td>
<td>Ukraine</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>25</td>
<td>Cyprus*</td>
<td>Cyprus</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>26</td>
<td>KP6*#</td>
<td>Serbia</td>
<td><em>G. pallida</em></td>
<td>This study results</td>
<td>This study results</td>
</tr>
<tr>
<td>No.</td>
<td>Sample Code</td>
<td>Country</td>
<td>Species</td>
<td>Sequencing Status</td>
<td>Gene Reference</td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>---------</td>
<td>------------------</td>
<td>--------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>27</td>
<td>Idaho25*</td>
<td>Idaho USA</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Pa1*#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>SCR/this study</td>
<td>AM409005</td>
</tr>
<tr>
<td>29</td>
<td>NiP3A*</td>
<td>S. America</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>SCR unpublished</td>
</tr>
<tr>
<td>30</td>
<td>P4A*#</td>
<td>S. America</td>
<td><em>G. pallida</em></td>
<td>SCR/this work</td>
<td>SCR unpublished</td>
</tr>
<tr>
<td>31</td>
<td>P5A*#</td>
<td>S. America</td>
<td><em>G. pallida</em></td>
<td>SCR unpublished</td>
<td>SCR unpublished</td>
</tr>
<tr>
<td>32</td>
<td>NiP4A*</td>
<td>S. America</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>SCR unpublished</td>
</tr>
<tr>
<td>33</td>
<td>NiP5A*</td>
<td>S. America</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>This study results</td>
</tr>
<tr>
<td>34</td>
<td>Otzuco*</td>
<td>S. America</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>AY851624.1</td>
</tr>
<tr>
<td>35</td>
<td>Luffness Field2#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>This study results</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>36</td>
<td>Luffness Field3</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>37</td>
<td>Luffness Field5</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>38</td>
<td>Luffness 09Field#</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>This study results</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>39</td>
<td>Ro1#</td>
<td>Unknown origin</td>
<td><em>G. rostochiensis</em></td>
<td>Fail to sequence</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>40</td>
<td>Ro2#</td>
<td>Unknown origin</td>
<td><em>G. rostochiensis</em></td>
<td>This study results</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>41</td>
<td>Ro3#</td>
<td>Unknown origin</td>
<td><em>G. rostochiensis</em></td>
<td>This study results</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>42</td>
<td>Ro4#</td>
<td>Unknown origin</td>
<td><em>G. rostochiensis</em></td>
<td>This study results</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>43</td>
<td>Ro5 Harmetz#</td>
<td>Unknown origin</td>
<td><em>G. rostochiensis</em></td>
<td>This study results</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>44</td>
<td>Ponikve#</td>
<td>Serbia</td>
<td><em>G. rostochiensis</em></td>
<td>Fail to sequence</td>
<td>Not sequenced</td>
</tr>
<tr>
<td>45</td>
<td>Muirhouse97</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Fail to sequence</td>
</tr>
<tr>
<td>46</td>
<td>Muirhouse88 (SASA)</td>
<td>Scotland</td>
<td><em>G. pallida</em></td>
<td>Not sequenced</td>
<td>Fail to sequence</td>
</tr>
</tbody>
</table>

*Included in CytB study
#Included in sequence analysis study of s222 mtDNA region

MOLECULAR STUDIES

DNA EXTRACTION

Genomic DNA was extracted from single second stage juveniles using MicroLysis Plus (Cambio, Cambridge, UK). A single cyst was placed in a 2.0 ml Eppendorf tubes with 1 ml of freshly made potato root exudate. After J2s hatched, each J2 was separately transferred onto a microscopic plate in 5 μl MicroLysis solution, where they were cut using a sterile surgical scalpel blade and then transferred to a 200 μl PCR Eppendorf tube. The slide was washed with 15 μl MicroLysis solution, this was combined with the other extract and then the samples were processed following the manufacturer’s protocol.

Genomic DNA was also extracted from 50 randomly taken cysts using the Qiagen Blood and Tissue Kit (Qiagen, Crawley, West Sussex, UK). Cysts were crushed in 180 μl ATL buffer using a plastic pellet homogenizer and a 1.5 ml Eppendorf tubes for 5 minutes. Then 18 μl proteinase K was added from the kit and samples incubated at 56°C for 2 h. Samples were
then centrifuged for 1min at 8000G, the supernatant transferred to a new tube and then processed according to the manufacturer’s instructions. DNA was eluted with 50 µl elution buffer, twice.

POLYMERASE CHAIN REACTION

All DNA extracts were tested with a diagnostic multiplex PCR assay for distinguishing PCN species (Ferris et al., 1993, Bulman and Marshall, 1997) to determine species identification of the samples before being used for mtDNA PCR. PCR amplification reactions were carried out in 10 µl reactions comprised of 0.04 µl (0.2 units) Taq (Promega, Southampton, UK), 0.4 µl of 10 µM of each primer and 1 µl DNA, 2 µl 5× buffer, 1 µl 0.2 mM dNTPs, 0.6 µl 2.5 mM MgCl2 and 4.16 µl HPLC water. Amplification conditions were an initial denaturation step at 94°C for 5 min followed by 40 cycles of 94°C for 30 s, 55°C for 15 s and 72°C for 30 s. A final extension cycle on 72°C for 5 min was used to terminate the reaction. The PCR program was performed using a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems, Warrington UK). The amplified products were separated by electrophoresis in 1% agarose gels buffered with 1%Tris-Borate-EDTA (TBE) (Sambrook et al., 2007) with Syber-safe dye, and visualised with UV illumination.

The CytB gene and part of the noncoding region on scmt II and scmt IV were carried out using PureTaq Ready-To-Go PCR Beads (GE Healthcare UK Ltd, Little Chalfont, UK) in 25 µl PCR comprised of 21 µl HPLC water, 1 µl of each primer and 2 µl of template DNA. Amplification conditions were as described above except with an extension step of 72°C for 1 min. Two specific primers INRACytbR and INRACytbL(Table 2) were used to amplify CytB (Picard et al., 2007) for 24 of total 34 populations for which sequence did not exist in Gene Bank and the SCRI database. Part of the noncoding region of scmtDNAs II and IV (s222)
which share some homology to the part of the longer region, x222 previously described by Armstrong et al., (1998) were amplified simultaneously in 50 µl reactions using primers F3mtDNA 222 and SCMT4-8 (Table 2). For T-RFLP analyses, F3mtDNA 222 and SCMT4-8 were fluorescently at the 5’ end with FAM and HEX respectively and used in amplification reactions as described above.

**Table 2.** List of the primers used in this study, their sequences, use and sources.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer Sequence</th>
<th>Use of primers</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>PITSr3</td>
<td>5’-AGCGCAGACATGGCCGCAA-3’</td>
<td>Distinguishing PCN species</td>
<td>Bulman and Marshall, 1997</td>
</tr>
<tr>
<td>PITSp4</td>
<td>5’-ACAACAGCAATCGTCGAG-3’</td>
<td>Distinguishing PCN species</td>
<td>Bulman and Marshall, 1997</td>
</tr>
<tr>
<td>PITS UNI</td>
<td>5’-CGTAACAAGGCTAGCTGTA-3’</td>
<td>Distinguishing PCN species</td>
<td>Ferris et al., 1993</td>
</tr>
<tr>
<td>INRAcytbL</td>
<td>5’-GGGTGTGGGGCTTATTTAC-3’</td>
<td>CytB gene amplification</td>
<td>Picard et al., 2007</td>
</tr>
<tr>
<td>INRAcytbR</td>
<td>5’-ACCAGCTAAAAACCCATCTC-3’</td>
<td>CytB gene amplification</td>
<td>Picard et al., 2007</td>
</tr>
<tr>
<td>F3mtDNA 222</td>
<td>5’-ATTAGACCAGATAAGTTACACCTTG-3’</td>
<td>S222 noncoding region</td>
<td>This study</td>
</tr>
<tr>
<td>SCMT4-8</td>
<td>5’-GACTAGGTCCATCAATCTGAAACC-3’</td>
<td>S222 noncoding region</td>
<td>This study</td>
</tr>
</tbody>
</table>

**SEQUENCING**

CytB PCR products were cut from the agarose gel using a scalpel and purified using a Qiagen MinElute Gel Extraction Kit. The DNA concentration of eluted DNA was determined using a NanoDrop spectrophotometer (Thermo scientific, Wilmington, USA), appropriately diluted, and sequenced in SCRI sequencing facility in both directions using primers INRAcytbR and INRAcytbL.

The s222 PCR products from single juveniles were gel purified as described above and cloned. Purified s222 DNA fragments were ligated in pGem-T Easy cloning vector (Promega, Southampton, UK) using 0.5 µl T4 DNA ligase, 0.5 µl of vector, 2.5 µl of ligation buffer and 1.5 µl of PCR product. Ligations were incubated at 15°C overnight. Electro-competent
dHP5α *Escherichia coli* cells (Invitrogen Ltd. Paisley, UK) were transformed and incubated on ampicillin-IPTG-X-gal LB agarose plates according to the manufacturer’s instructions. White colonies were further checked by PCR using s222 primers. Scrapes of each colony taken and added to 100 µl H20 in a 0.5 µl Eppendorf tube. Bacterial cells were burst at 95°C for 5 min and after cooling 1 µl used in a PCR reaction to check that the plasmid carried the insert. PCR amplification reactions were carried out in 10 µl reactions comprised of 0.04 µl (0.2 units) Taq (Promega, Southampton, UK), 0.4 µl of 10 µM of each primer and 1 µl burst cells, 2 µl 5× buffer, 1 µl 0.2 mM dNTPs, 0.6 µl 2.5 mM MgCl2 and 4.56 µl HPLC water. Amplification conditions were the same as for amplification of s222 from genomic DNA described previously. Colonies that were positive for the insert were grown O/N in LB containing ampicillin and the plasmid was purified using a GeneJetPlasmid Purification Kit (Fermentas, York, UK). Up to eight s222 clones for each J2 were selected for the 14 representative populations (Luffness09, Luffness99, LuffnessField2, Gourdie, Lindley, KP6Serbia, Pa1, P4A, P5A, Ro1, Ro2, Ro3, Ro4 and Ro5). Sequencing reactions were carried out in SCRI Sequencing laboratory using M13 forward and reverse primers (Sambrook *et al.*, 2007).

**PCR PRODUCT DIGESTION WITH RESTRICTION ENZYMES, RFLP AND T-RFLP ANALYSES**

15 µl of s222 PCR product were digested separately with TaqI, AluI and XbaI restriction enzymes according to the manufacturer’s instructions (Promega) with 5 µl enzyme mix, composed of 1 µl (10 units) enzyme, 2 µl appropriate buffer and 3 µl of HPLC water, for 3 hours at 37°C. Loading dye was added to 15 µl of the digestion solution and loaded in a 2% agarose gel to obtain an RFLP image and 5 µl was further processed for T-RFLP analysis.
DNA fragments were separated by electrophoresis in a 2% TBE buffered agarose gel and visualized using UV illumination.

For preparation of samples for T-RFLP analysis, following PCR and restriction digestion as described above, PCR the digest were diluted 1:10 in H20 and then for each sample 1 µl of diluted PCR digest were further diluted with 8.95 µl of formamide and 0.05 µl GeneScan™ ROX 1000 marker (Applied Biosystems, Foster City, USA) was added. Samples were analyses with an automated DNA sequencer (ABI PRISM™ 3730) and included undigested and negative control (H20) samples.

DATA ANALYSES

BIOINFORMATIC ANALYSIS

Sequences were edited and contigs of forward and reverse sequences produced using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, USA). Sequences were aligned with BioEdit using ClustalW with default parameters (Hall, 1999) then uploaded on http://www.phylogeny.fr/ (Dereeper et al, 2008) where they were further analyzed by default parameters, which includes realignment using MUSCLE (Edgar R.C., 2004), curation with selection of conserved blocks from multiple alignments (Gblocks, Castresana, J. 2000), phylogenetic estimation by maximum likelihood method (Phyml, Guindon, S. and Gascuel, O., 2003) and tree rendering by TreeDyn.

VIRTUAL DIGESTION OF S222

Virtual digestion of s222 sequences was performed for restriction enzymes TaqI, AluI and Xbal using http://insilico.ehu.es/restriction/main/ with default settings. Restriction patterns were checked and visualised with software GENtle. http://gentle.magnusmanske.de/.
ANALYSIS OF T-RFLP DATA

Genotyping data for the T-RFLP samples of the s222 PCR products digested with restriction enzymes, TaqI, AluI and XbaI were processed using GeneMapper (Applied Biosystems, Foster City, USA) to allow peak sizing and generation of a peak area for each identified peak. A fixed bin width of 5 bp was used with a threshold of 400 to reduce background. Two data sets were generated, one for terminal amplicons from the 5’ end fluorescently labelled with FAM (blue) and one data set for terminal PCR fragments from the 3’ end fluorescently labelled with HEX (green). Further analysis of these data sets was performed with EXCEL (Microsoft Corporation, Redmond, USA) to exclude fragment sizes below 132 and above 760 and fragment sizes for which there was no data. The resulting peak relative abundance data was then analysed with in Genstat 12th edition (VSN International Ltd.). Data were analyzed using the multivariate Principal Components Analysis (PCA) based on sums and squares and products, the default option. Charts and graphs were made in Genstat and Microsoft Excel.

Results

SEQUENCING

CYTB SEQUENCES

CytB gene sequences, 872bp in length, were obtained for 17 G. pallida populations. Sequencing failed for populations Bankhead of Buchel, Bedale, Muirhouse, Uzhok, Windyedge and Idaho25. Sequences are aligned together with others already deposited in GenBank, sequenced by Plantard et al., (2008) and Pylypenko et al., (2008) or SCRI unpublished sequences, using sequences of G. mexicana as the outgroup (Picard et al, 2007).
Sequence alignment and phylogenetic analysis resulted in a ML tree shown in Fig. 1. Plantard et al. (2008) reported that all European populations share the same ancestor, and our data is consistent with this proposition. All European populations formed a significantly supported clade with South American populations clustered in sister clades. The clade with European populations branches into three subclades, the largest of which contains a subgroup with the two Scottish populations Gourdie and Luffness09, a second with Pa1, and the third with all other European populations. Among this latter group the Serbian, German population Kalle, Zhornava from Ukraine, and Cyprus populations cluster with two populations from The Netherlands.

*Figure 1.* Phylogenetic tree (left) and part of the alignment (right) showing the relationships between *Globodera pallida* populations based on partial Cytochrome B sequences. Additional sequences obtained from the present study are underlined.

**S222 SEQUENCES**

Sequences were obtained from clones of the s222 PCR product from cysts of *G. pallida* populations (6 for Muirhouse95 and 2 for Bankhead of Buchel), and single juveniles from *G. pallida* (Luffness09, Luffness09field, LuffnessField2, Gourdie, Lindley, Pa1, P4A and KP6) and *G. rostochiensis* populations (Ro2, Ro3, Ro4 and Ro5). Edited sequences were aligned
with unpublished sequences from SCRI and those already in the database (Table 1). Sequences obtained from single juveniles presented in Table 3.

The s222 sequences formed two groups when aligned with representative sequences of scmtDNAIV (LindleyDQ631913, P4ADQ28831, P4ADQ28833 and P4ADQ288936), scmtDNAII (LindleyDQ631911) and other sequences from scmtDNAIV (Fig. 2). Six groups were formed with the same type of sequences, 3 representing scmtDNAII (scmtDNAII Type 1, 2 and 3) and 3 other representing scmtDNAIV (scmtDNAII Type 1, 2 and 3) G. rostochiensis sequences forming the outgroup. ScmtDNAII Types clustered together with support value of 0.12 with a sister group scmtDNAIV Type 1, forming a well supported clade. This supergroup is related to Type 2 and Type 3 from scmtDNAIV. The vast majority of sequences from South American populations P4A and P5A were found in scmtDNAIV Type 1, with only 4 of 26 sequences placed in two other Types. The Luffness populations were seen in all these Types except the scmtDNAIV Type 1. The scmtDNAIV Type 3 was comprised mostly of Luffness sequences, with the exception of one P4A sequence. All Pa1 sequences, together with all Muirhouse and 3 P4A sequences comprised the scmtDNAII Type 1. ScmtDNA IV Type 2 consists mostly of Lindley, Gourdie and KP6 sequences, where six sequences of KP6 formed a well supported subgroup. The same sequence types are shared among different populations, which has was previously been reported.
Figure 2. A maximum-likelihood phylogenetic tree representing the relationships between *Globodera pallida* populations based on s222 sequences from scmtDNAII and scmtDNAIV and *G. rostochiensis* s222 sequences. Additional sequences obtained from the present study are underlined.

Sequences obtained form juveniles of populations Luffness09, Gourdie and KP6 appear to have 2 sequence types among scmtDNAII, where scmtDNAIV shows less sequence diversity with one type in all populations, as presented in Table 3. This is the first observation of more than one form of scmtDNA II within individual nematodes. For example, the clones from single juveniles from Luffness09 (Luff093.3, Luff093.6, Luff093.8, Luff093.1, Luff09j23-2), Gourdie (GourdieJ2.3, GourdieJ2.2) and Serbia (KP6J2.7, KP6J2.3) are presented as different types in the alignment and the phylogenetic tree in Fig. 3.

Table 3. Frequency of different sequence types of scmtDNAII and scmtDNAIV in clones of s222 from single juveniles from populations of *Globodera pallida* and number of s222 *G. rostochiensis* sequences.

<table>
<thead>
<tr>
<th>Juveniles</th>
<th>scmtDNAII type1</th>
<th>scmtDNAII type2</th>
<th>scmtDNAII type3</th>
<th>scmtDNAIV type1</th>
<th>scmtDNAIV type2</th>
<th>scmtDNAIV type3</th>
<th>total # clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>LindleyJ2-2</td>
<td></td>
<td></td>
<td></td>
<td>1,2,3,6,7,8,9</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>GourdieJ2-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>GourdieJ2-2</td>
<td></td>
<td>2</td>
<td>3</td>
<td></td>
<td>1,5,6,7,9</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>KP6J2-2</td>
<td></td>
<td>3</td>
<td>7</td>
<td></td>
<td>1,2,4,5,6,9</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Luffness09J2-1</td>
<td></td>
<td></td>
<td></td>
<td>2,4,6</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Luffness09J2-3</td>
<td></td>
<td>1,2</td>
<td>3,8</td>
<td></td>
<td>1,4,5,7</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Luffness09J2-4</td>
<td></td>
<td>1,4</td>
<td></td>
<td></td>
<td>2,3,5</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Luffness09J2-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5,7</td>
<td>2</td>
</tr>
<tr>
<td>Luffness99J2-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,5,6</td>
<td>5</td>
</tr>
<tr>
<td>LuffnessField2*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5,6</td>
<td>4</td>
</tr>
<tr>
<td>Pa1J2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,2</td>
<td>8</td>
</tr>
<tr>
<td>P4AJ2</td>
<td>1,2,3,4,5,6,7,8</td>
<td></td>
<td></td>
<td></td>
<td>1,3,4,5,6,7,8,9</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Ro2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ro3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ro4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ro5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78</td>
</tr>
</tbody>
</table>
Figure 3. Alignment of scmtDNAII s222 sequences (left) and phylogenetic tree with the same sequences (right) of s222 sequences from single juveniles of *G. pallida*

VIRTUAL DIGESTION OF S222

Virtual digestion all s222 sequences with restriction enzymes TaqI, AluI and XbaI showed that TaqI cut three times with most of the sequences, AluI had two restriction sites in almost all sequences, and XbaI digested less than third of sequences one time. Positions of restriction sites and sequence length values were used for calculating lengths of all theoretically possible cut fragments, both in the case of complete digestion and possible uncut or partially cut fragments, named as 5’A, 5’B, 5’C, 5’D from the 5’end, and 3’A, 3’B, 3’C, 3’D from the 3’ end of the s222 sequence types (supplementary data).

By combining the digestion patterns for all three enzymes generated for four representative sequence types of scmtII and IV, all of the types can be differentiated. Digestion patterns for every enzyme were different for the scmtII and scmtIV types. Figure 4 shows the differentiation of four sequence types based on terminal fragment lengths (5’A and 3’A) only. Sequences that formed the first two groups (Type 11 and Type12), are the same sequences
which clustered together forming group 1 in scmtDNAII and scmtDNAIV, and the third and fourth sequence groups (type2 and type3) are consistent with groups 2 and 3 on scmtDNAII and scmtDNAIV in Fig. 2.

![ScmtDNAII terminal fragments](image_url)

<table>
<thead>
<tr>
<th>Type</th>
<th>5' A</th>
<th>3' A</th>
<th>5' A</th>
<th>3' A</th>
<th>5' A</th>
<th>3' A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type1</td>
<td>297</td>
<td>637</td>
<td>704</td>
<td>230</td>
<td>934</td>
<td>934</td>
</tr>
<tr>
<td>Type2</td>
<td>304</td>
<td>318</td>
<td>704</td>
<td>230</td>
<td>231</td>
<td>704</td>
</tr>
<tr>
<td>Type3</td>
<td>308</td>
<td>318</td>
<td>203</td>
<td>225</td>
<td>940</td>
<td>940</td>
</tr>
</tbody>
</table>

![ScmtDNAIV terminal fragments](image_url)

<table>
<thead>
<tr>
<th>Type</th>
<th>5' A</th>
<th>3' A</th>
<th>5' A</th>
<th>3' A</th>
<th>5' A</th>
<th>3' A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type1</td>
<td>167</td>
<td>625</td>
<td>567</td>
<td>226</td>
<td>794</td>
<td>794</td>
</tr>
<tr>
<td>Type2</td>
<td>298</td>
<td>355</td>
<td>730</td>
<td>239</td>
<td>969</td>
<td>969</td>
</tr>
<tr>
<td>Type3</td>
<td>311</td>
<td>555</td>
<td>694</td>
<td>265</td>
<td>384</td>
<td>575</td>
</tr>
</tbody>
</table>

**Figure 4** Bar charts and tables representing lengths of terminal fragments among different sequence types obtained by *in silico* digestion of s222 sequences from scmtDNAII and scmtDNAIV of *G. pallid*. 
RFLP ANALYSES

RFLP analysis with TaqI showed differentiation of 35 *G. pallida* populations into three main restriction pattern groups (Fig 5). One group includes the South American populations, the second includes most of the European populations and the third consists of Pa1. AluI and XbaI also distinguished the populations into the same groups (results not shown).

Figure 5 RFLP agarose gel image representing restriction enzyme digestion with TaqI. Order of samples as in Table 1.

However, among European populations, population Muirhouse95 and Luffness09 digested with TaqI showed different patterns compared to most other European populations. The pattern for Muirhouse95 was more similar to the restriction patterns formed by South American populations, identical to P3A, and Luffness09 was most similar to Pa1. To check the validity of these results, DNA was extracted again from the Muirhouse95 cysts from the same source and another SCRI stock of the Muirhouse97 population, and digestion of these s222 PCR products gave an identical pattern to the original Muirhouse95 DNA extraction (Fig 6B). However another Muirhouse population obtained from SASA, did not confirm these digestion patterns showing restriction pattern typical of European populations. The s222 region of Muirhouse95 was cloned and sequenced and sequences all clustered with scmtDNAII sequences from P4A and Pa1 sequences, and had an identical virtual restriction pattern as well (Table 4B).
The Luffness99 population which was previously recorded as being different molecularly and by its virulence (Blok, 1997), showed a restriction pattern similar to Scottish population Pa1. Three fields (2, 3 and 5) in the vicinity of the field where the original Luffness population was isolated were sampled in 2009 and 2010. Cysts were recovered and DNA was extracted. Digestion of the s222 PCR product with TaqI showed that two cysts from field 2 had a highly similar though not totally identical pattern as previously found for Luffness09. Cysts from field 3 and 5 had similar restriction patterns to that of typical European populations (Fig 6A). Cloning and sequencing of the s222 region from the first cyst from Field 2 failed however, sequencing of cyst 2 from this field was successful and four sequences were obtained. Two clones showed virtual restriction patterns like Pa1, Muirhouse95 and some of the P4A sequences which all were clustered in Type 1 representing one group of sequences from scmtDNAII (Table 4B). The other two sequences clustered with scmtDNAIV sequences, Type 3, where LuffField2.1 showed an identical restriction pattern with the group while sequence LuffField2.2 was more similar in restriction pattern with sequences from Type 2 (Table 4A). These results confirmed the existence of atypical *G. pallida* population in UK.
Table 4 Positions of restriction enzyme TaqI, AluI and XbaI digestion sites in clones of s222 PCR amplification products from Luffness, Luffnessfield, P4A, Pa1 and Muirhouse.

4A- Restriction sites in sequences from scmtDNAIV group 2
4B- Restriction sites in sequences from scmtDNAII group 1

<table>
<thead>
<tr>
<th>Population</th>
<th>Total sequence length</th>
<th>TaqI cut position</th>
<th>AluI cut position</th>
<th>XbaI cut position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1  2  3</td>
<td>1  2  3</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone4</td>
<td>958</td>
<td>310 403</td>
<td>693 383</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone5</td>
<td>958</td>
<td>310 403</td>
<td>693 383</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone8</td>
<td>958</td>
<td>310 403</td>
<td>693 383</td>
<td></td>
</tr>
<tr>
<td>Luff99J21.2</td>
<td>958</td>
<td>311 404</td>
<td>693 384</td>
<td></td>
</tr>
<tr>
<td>LuffField2.1</td>
<td>956</td>
<td>311 404</td>
<td>693 384</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone1</td>
<td>959</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
<tr>
<td>LuffField2.2</td>
<td>949</td>
<td>299 393</td>
<td>684</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone2</td>
<td>959</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone3</td>
<td>959</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone6</td>
<td>959</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
<tr>
<td>LuffscIVclone7</td>
<td>959</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
<tr>
<td>P4ADQ288935</td>
<td>960</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
<tr>
<td>Luff99J21.4</td>
<td>957</td>
<td>311 404</td>
<td>694 384</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population</th>
<th>Total sequence length</th>
<th>TaqI cut position</th>
<th>AluI cut position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1  2  3</td>
<td>1  2</td>
</tr>
<tr>
<td>Muirhouse95.2</td>
<td>934</td>
<td>296 391</td>
<td>704</td>
</tr>
<tr>
<td>Pa1J2-7</td>
<td>934</td>
<td>296 391</td>
<td>704</td>
</tr>
<tr>
<td>Muirhouse95.1</td>
<td>933</td>
<td>297 392</td>
<td>703</td>
</tr>
<tr>
<td>Muirhouse95.3</td>
<td>934</td>
<td>297 392</td>
<td>704</td>
</tr>
<tr>
<td>Muirhouse95.4</td>
<td>934</td>
<td>297 392</td>
<td>704</td>
</tr>
<tr>
<td>Muirhouse95.5</td>
<td>933</td>
<td>297 391</td>
<td>704</td>
</tr>
<tr>
<td>Muirhouse95.6</td>
<td>936</td>
<td>298 393</td>
<td>706</td>
</tr>
<tr>
<td>P4ADQ288931</td>
<td>933</td>
<td>297 391</td>
<td>613 703</td>
</tr>
<tr>
<td>P4ADQ288936</td>
<td>934</td>
<td>297 392</td>
<td>614 704</td>
</tr>
<tr>
<td>Pa1J2-1</td>
<td>935</td>
<td>297 392</td>
<td>615 705</td>
</tr>
<tr>
<td>Pa1J2-2</td>
<td>932</td>
<td>297 392</td>
<td>615 705</td>
</tr>
<tr>
<td>Pa1J2-3</td>
<td>935</td>
<td>297 391</td>
<td>613 702</td>
</tr>
<tr>
<td>Pa1J2-5</td>
<td>935</td>
<td>297 392</td>
<td>615 705</td>
</tr>
<tr>
<td>Pa1J2-6</td>
<td>935</td>
<td>297 392</td>
<td>615 705</td>
</tr>
<tr>
<td>P4ADQ288933</td>
<td>935</td>
<td>298 393</td>
<td>615 705</td>
</tr>
<tr>
<td>Pa1J2-4</td>
<td>935</td>
<td>298 393</td>
<td>615 705</td>
</tr>
<tr>
<td>Pa1J2-8</td>
<td>935</td>
<td>298 393</td>
<td>616 705</td>
</tr>
<tr>
<td>LuffField2.5</td>
<td>934</td>
<td>298 392</td>
<td>614 704</td>
</tr>
<tr>
<td>LuffField2.6</td>
<td>933</td>
<td>298 392</td>
<td>614 704</td>
</tr>
</tbody>
</table>
T-RFLP

T-RFLP analysis gave similar results to those obtained with RFLP for TaqI and AluI. Using these enzymes, T-RFLP clearly differentiated the South American group together with the atypical European populations mentioned above from the remaining majority European group with common characteristics as presented by previous techniques. However, the enzyme XbaI failed to give this distinction. The PCA analysis accounted for 44.07 – 71.02% of the variation in the first and 16.48 – 30.34% in second PCA axis. The results for TaqI and AluI were consistent when analysed separately (Figs 7A and B) and when all three data sets were analysed together (Fig 7C). In all scatter plots population Luffness09 was separated from all other sequences. South American populations were always clustered with different distances among each other but always clearly apart from the other European populations. Within this last group some of the European populations previously classified as typical were separated from the main European group. For example population Windyedge separates in PCA analysis using Taq green and TaqI complete data set. In the same manner population D375 from The Netherlands appeared similar to South American population Otzuco using data set AluI green.
By looking at latent vectors values and their influence on data distribution we found that among a total of 40 peak sizes registered from our samples by T-RFLP, 20 of them have influence on the data distribution of less than 5%. The majority of peaks which have strong influence recorded by T-RFLP are the same or close to the values obtained performing an *in silico* digest. Among 15 terminal fragment peaks that influenced results with values more than 10% obtained by T-RFLP, 5 with the exact length and 5 fragments with 2-3 nucleotides difference in length (Fig 8A) were found in data set done by virtual digestion (Fig 8B and C). Interestingly, among 4 fragments with highest influence above 40%, two of them were exact matches (167 and 544) and one, fragment 624, differed in length in 2 or 3 nucleotides (Fig. 8A).
Figure 8 Terminal fragment analyses from T-RFLP automated sequence analysis and \textit{in silico} digestion of s222 amplification products with TaqI. A) Histogram showing latent vector loadings of TaqI digested s222 T-RFLP terminal fragments sizes (x axis) labelled with HEX-SCMT4-8 5’ (green) and FAM-F3mtDNA 222 (blue) and their influence on data distribution for PCA axis 1 and 2, B) and C) 5’ and 3’ fragment sizes from \textit{in silico} analyses of s222 scmtII and scmtIV sequence types.

Discussion
CYTB ANALYSES

The relationships of *G. pallida* populations found in this study are consistent with results reported by Plantard *et al.*, (2008) and Pylypenko *et al.*, (2008). All European populations were grouped into clade I as previously reported. These results support the proposition that the *G. pallida* populations found in Europe are from a limited geographic region in South America in southern Peru. There is also evidence to suggest that there are different variants of European *G. pallida* populations within this clade, which could indicate that more than one introduction has occurred (Plantard *et al.* (2008)). Clustering of recently recorded populations from Serbia, Ukraine and Cyprus, together with populations Kalle from German, and D375 and GP-74 from the Netherlands suggest that these populations of *G. pallida* are related and may have originated from the Netherlands, one of the biggest seed potato exporters in Europe.

S222 SEQUENCE ANALYSES

S222 noncoding sequence of mtDNA of *G. pallida* is highly polymorphic and has sufficient variation for separating several groups populations, including those in different pathotypes. The sequences of the s222 region from scmtII and IV obtained in this study combined with published and unpublished sequences showed that different sequence types are shared among populations in agreement with the results of Armstrong *et al.*, (2007). There were at least six main sequence types among the two scmtDNAs. Populations differed in the number of clones of either scmtII or scmtIV suggesting that populations may differ in the representation of these two scmtDNAs in their complement of scmtDNAs, however, the number of clones sequenced is relatively small so further sequencing or quantitative PCR would be required to confirm this. While the amplification of scmtII and IV in the same PCR reaction produces a complex mixture of PCR products due to the lack of specificity of the primers for either scmtDNA and the potential presence of more than one scmt type within a populations, this
also increases the information obtained in each sample and the potential to differentiate it from other populations. Further analyses are needed to assess whether specific primers for one or other of the two scmtDNAs examined in this study would be sufficient to provide optimal intra-specific differentiate of *G. pallida* populations.

Some populations showed more sequence types than others, such as population Luffness09, Gourdie, KP6 and P4A with sequences present in three different groups, compared to Pa1 and the Muirhouse95 population with just one sequence type of scmtDNA II represented. Sequences obtained from juveniles showed more diversity compared DNA extracted from multiple cysts where almost all sequences were of the same type for almost all populations. Possible reasons could be predominance of certain type of sequences among each population which was preferentially amplified due to larger concentration in the mtDNA population.

The Pa1 pathotype was distinguished from the main group of European *G. pallida* populations in the CytB sequence analyses however, s222 sequences from from Pa1 grouped with those from P4A, Muirhouse and Luffness (Fig 2). On further investigation of the Muirhouse95 and 97 samples it was found that they differed from the Muirhouse sample obtained from SASA suggesting that a mix up has occurred with this population at some point. The P4A sequences in the scmtII group with that of Pa1 were annotated prior to the sequencing of scmtDNA II (Gibson *et al.*, 2007) so these sequences probably need to be re-annotated as scmtDNAII. Interestingly there were no scmtDNA IV sequences obtained for Pa1 suggesting that this scmt may be absent or infrequent in this population confirming the results of Armstrong *et al.* (2000).
The Luffness samples comprised the original unusual isolate, and recent samples from the Luffness farm taken in 2009 and 2010. The sample obtained in 2009 from field 3 was like the typical European Pa2/3 type whereas the 2010 sample from Field 2 was similar to the original isolation. Field 2 was closer to the field where the original sample was found than field 3. This indicates that it is possible to differentiate field populations from the same farm and presents the first demonstration of the use of a molecular method that has the potential to distinguish populations with different virulence characteristics.

Amplification products were obtained with the F3mtDNA 222 and SCMT4-8 primers from single juveniles of *G. rostochiensis* Ro2, 3, 4 and 5. The sequences of these products were shorter than those for *G. pallida*, highly A and T rich and showed limited homology to *G. pallida* sequences. Confirmation that these sequences are of mtDNA origin is needed an further analyses of more *G. rostochiensis* populations is needed to assess the intra-specific relationships in this species.

**RFLP AND T-RFLP ANALYSES**

The RFLP and T-RFLP analyses in this study demonstrate the use of two different techniques for the analysis of intra-specific separation of *G. pallida* populations. The regions targeted for the analyses are based on highly polymorphic regions of *G. pallida* mitochondrial DNA. PCR primers were used that were designed to amplify a region of ~1kb from the scmtDNAII and scmtDNAIV for *G. pallida* base on published sequences derived from populations Lindley and Luffness. For the RFLP analyses, PCR amplification products were digested with a restriction enzyme (TaqI, AluI or XbaI) and then subjected to agarose gel electrophoresis. The enzyme TaqI readily distinguished 3 groups within *G. pallida* as did the other two enzymes. The three groups contain populations which are recognised as different pathotypes
(Pa1, Pa2/3 and P4A and P5A) (Phillips and Trudgill, 1998). The equipment required for the RFLP analyses is found in many plant pathology laboratories that undertake PCR based diagnostics and thus the method described could be used to check *G. pallida* populations for unusual mtDNA patterns. Restriction pattern uniformity among sequences in each group suggests that this variable region is conserved enough to be used for population studies.

This was the first attempt to use T-RFLP analyses for a population genetics study of a nematode species to our knowledge. The aim was to find a sensitive and high-throughput technique which could distinguish different groups of *G. pallida* populations and possible new introductions of *G. pallida*. T-RFLP analysis has the potential for analysis and differentiation of distinct population groups in a large scale (George *et al*., 2009), in a relatively quick and inexpensive way, compared with traditional sequencing. Both the CytB sequences and s222 sequences were able to differentiate several groups within *G. pallida* however the RFLP method does not require sequencing of the amplification product. The T-RFLP method may be useful for high through-put analysis of PCN populations and further investigation may provide a more sensitive method to differentiate populations within the main groups. Combined with other DNA based diagnostic methods for species determination, either RFLP or T-RFLP analyses can provide additional information about the intra-specific characteristics of *G. pallida* populations. The occasional anomalies in the results suggests that further optimisation of the method is needed this approach has potential for early detection of unusual populations of *G pallida* which may have unusual virulence characteristics. This could be of interest for institutions that are examining huge number of samples such as in quarantine services. However the technique should be improved by finding optimal dilutions of PCR products for example, and optimising appropriate parameters for analysing the data sets.
Acknowledgements

On the first place I want to thank Vivian Blok on her patience and constructive and supportive guidance while supervised this work. I acknowledge Tim Daniell, Susan Mitchell for advices and assistance with T-RFLP analysis and Mark Phillips for useful comments. I thank Luffness Loan Farm for providing me with field samples and Dr Jon Pickup and Tom Archibald at SASA for PCN samples. This work was supported by EUMAINE Erasmus Mundus grant.
References


